Accepted refereed manuscript of: Bartie KL, Taslima K, Bekaert M, Wehner S, Syaifudin M, Taggart JB, de Verdal H, Rosario W, Muyalde N, Benzie JAH, McAndrew BJ & Penman DJ (2020) Species composition in the Molobicus hybrid tilapia strain. Aquaculture, 526, Art. No.: 735433. https://doi.org/10.1016/j.aquaculture.2020.735433

Research Article

Species composition in the Molobicus hybrid tilapia strain

Kerry L. Bartie[a,1], Khanam Taslima[a,b,1], Michaël Bekaert[a*], Stefanie Wehner[a], Mochamad Syaifudin[a,c], John B. Taggart[a], Hugues de Verdal[d,e,f], Westly Rosario[g], Nerafe Muyalde[g], John A.H. Benzie[d,h], Brendan J. McAndrew[a], David J. Penman[a]

[a] Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK

[b] Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh

[c] Program Study of Aquaculture, Agriculture Faculty, Sriwijaya University, South Sumatra, Indonesia

[d] WorldFish Centre, Jalan Batu Maung, 11960 Bayan Lepas, Penang, Malaysia

[e] CIRAD, UMR ISEM, 34398 Montpellier, France

[f] ISEM, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France

[g] BFAR-NIFTDC, Bonuan-Binloc, Dagupan City, Philippines

[h] School of Biological Earth and environmental Sciences, University College Cork, Cork, Ireland

[*] E-mail: michael.bekaert@stir.ac.uk

[1] These authors contributed equally to this work

Keywords: Tilapia; GIFT; Molobicus; selective breeding; species-specific SNP markers; aquaculture

25

Research Highlights

- Novel SNP markers revealed species contribution within a hybrid tilapia line (Molobicus) undergoing selection for growth performance in brackish water.
- The backcross base population, developed from feral *Oreochromis mossambicus* and GIFT, matched the predicted 3:1 ratio for *O. mossambicus* and *O. niloticus* respectively.
- The selected lines showed a significant increase in *O. niloticus*-specific alleles.

Abstract

The “Molobicus” hybrid breeding programme was initiated to improve tilapia growth performance in brackish water. The base population was created by backcrossing F_1 *Oreochromis niloticus* GIFT strains × feral *O. mossambicus* to *O. mossambicus* and selective breeding conducted for performance in brackish water with two lines selected in extensive or intensive rearing conditions. A panel of ten diagnostic SNP markers was applied to estimate the species composition at different stages in the Molobicus programme including parental stocks, F_1 and seventh generation fish from the selective lines. The *O. aureus*-specific markers tested revealed zero or negligible contribution from *O. aureus* to all the groups analysed. Feral *O. mossambicus* possessed an estimated 0.98 frequency of *O. mossambicus*-specific alleles, while GIFT samples had an estimated mean frequency of 0.88 *O. niloticus*-specific alleles. Hybrid F_1 GIFT × feral *O. mossambicus* samples demonstrated close to 50:50 allele frequencies from *O. niloticus* and *O. mossambicus* for seven of the eight SNP loci tested. Analysis of the combined seventh generation Molobicus samples revealed a significant excess of *O. niloticus* alleles in six out of the eight SNPs tested, with this trend being more pronounced in the line selected in intensive culture conditions and showing increased body weight. PCR-based SNP assays such as these can be used to inform on the individual species contribution of fish stocks and provide tools for the genetic management of the tilapia species and future breeding programmes.

1. Introduction

Aquaculture production of tilapias (family Cichlidae), native to Africa and the Middle East, is currently the second highest of any finfish group globally after carps (FAO, 2019). Three species belonging to the *Oreochromis* genus predominate: *O. niloticus* (Nile tilapia), *O. mossambicus* (Mozambique tilapia), and *O. aureus* (Blue tilapia). In the second half of the last century, these fish were widely distributed in Asia and other tropical and semi-tropical regions to increase food production due to commercially desirable traits such as high growth rates, ability to survive in different aquatic environments and resistance to disease (Modadugu and Acosta, 2004).

Following the introduction of tilapia species outside their native ranges, concern was raised regarding the genetic management and conservation of the farmed broodstock due to introgression (Macaranas et al., 1986) and low effective population sizes (Pullin and Capili, 1988), resulting in poor performance (Eknath et al., 1991). There was therefore a drive to initiate selective breeding programmes (Gjedrem et al., 2012). The Genetic Improvement of Farmed Tilapia (GIFT) project, started in 1988 by WorldFish (then ICLARM) and partners, was the first major breeding programme designed to improve the performance and supply of high-quality *O. niloticus* stocks suitable for both small-scale and commercial aquaculture systems in Asia (Eknath et al., 1998).

To ensure a broad genetic diversity, the base *O. niloticus* populations for GIFT were sampled from wild stocks from Africa and farmed tilapia stocks in Asia and Israel. However, the exact species makeup of the founder populations used in GIFT and other tilapia breeding programmes is unknown, due to the likelihood of introgression into the farmed stocks used (Angienda et al., 2011; Firmat et al., 2013; Neira, 2010).

Hybridisation between tilapia species has in some cases been intentional to promote desirable traits in aquaculture, as seen in the production of F_1 hybrids using *O. niloticus* (favoured for rapid growth) and *O. aureus* (tolerant to colder temperatures), also popular due to the high male percentage, that now contribute significantly to the total tilapia production in China under variable climatic

conditions (Cai et al., 2004). A breeding programme based on hybrid tilapia, called “SaltUNO” or Molobicus, began in the Philippines in 1999 to improve performance in brackish water. The base population was developed from seventh generation GIFT *O. niloticus* strains (selected for growth) and feral *O. mossambicus* (saline tolerant) captured from wild stocks in the Philippines (Mateo et al., 2004 ; Figure 1). The resulting F_1 hybrid (GIFT *O. niloticus* × feral *O. mossambicus* parent) was backcrossed with feral *O. mossambicus* to improve the salinity tolerance, and generations of selective breeding followed from this backcross, targeting increased growth performance (body weight at five months) and passive selection by rearing fish in brackish water (de Verdal et al., 2014). Two selected lines were developed, one selected in extensive culture conditions (fertilised earthen ponds, without additional feed and at low stocking density) and one in more intensive culture (tank based with *ad libitum* feed and at high stocking density).

Molecular differentiation of tilapia species is possible using protein-based allozyme loci, but this method requires destructive sampling and the number of informative markers is limited (Sodsuk and McAndrew, 1991). DNA-based markers provide a greater discrimination potential to measure genetic diversity. Methods such as random amplified polymorphic DNA (RAPD; Bardakci and Skibinski, 1994; Dinesh et al., 1996), microsatellite markers (Costa-Pierce, 2003) and PCR based restriction fragment length polymorphism (RFLP; Toniato et al., 2010) have been used to characterise tilapia species, but none of these techniques give sufficient informative markers to confirm the species purity of individual fish or to assess levels of introgression. Mitochondrial DNA (mtDNA) sequencing has also be applied to separate tilapia species (D’Amato et al., 2007; Wu and Yang, 2012) but this is of limited use for studies of hybridisation and introgression as mtDNA is maternally inherited.

The advent of high throughput sequencing has allowed the identification of thousands of single nucleotide polymorphism (SNP) markers (Kumar et al., 2012) to assess genetic diversity within populations, differentiate between species and subspecies and map loci associated traits (Palaiokostas et al., 2013; Van Bers et al., 2012; Xia et al., 2014). One such sequencing technique, restriction-site

associated DNA sequencing (RADseq), offers a reduced representation of the genome and is able to generate SNP markers randomly distributed throughout the genome adjacent to restriction enzyme cut sites (Davey and Blaxter, 2010). A recent study using a double digest variant of RADseq (Peterson et al., 2012) , identified species-specific SNPs for ten different tilapiine species and validated 24 putative SNP markers for four species of tilapia commonly used in aquaculture using a PCR based SNP assay (Syaifudin et al., 2019).

The aim of the present study was to apply selected species-specific SNP markers for three tilapia species (*O. niloticus*, *O. mossambicus* and *O. aureus*) and to assess the species composition at various stages in the Molobicus breeding programme comprising parental feral *O. mossambicus* and GIFT strains, the F_1 hybrid and the seventh generation Molobicus hybrid fish selected in two farming systems. These results from the SNP markers provide insights on how the selective pressures present within the Molobicus breeding program shaped the species contribution and genomic profile of the selected hybrid lines.

2. Materials and Methods

2.1 Ethical Statement

Archived fin samples were obtained from the Molobicus and GIFT breeding programmes and approved for analysis at the University of Stirling by the University Animal Welfare and Ethical Review Body (AWERB).

2.2 Sample collection

Fin clip samples stored in 95% ethanol were obtained from fish involved the Molobicus breeding programme in the Philippines (de Verdal et al., 2014). These samples comprised parental stocks of feral *O. mossambicus* (n=23), F_1 hybrid samples (n=20) from the initial crossing between feral *O. mossambicus* and seventh generation GIFT strains, and 58 individuals (derived from 17 families) from the seventh generation (G7) of the Molobicus breeding programme, following selective breeding for increased body weight at five months (derived from the F_1 hybrid backcrossed with *O. mossambicus*). Within the selective breeding programme, there were two separate selected lines, reared in two different farm environments, either an extensive pond-based system (n=24 samples) or an intensive tank facility (n=34). An overview of the Molobicus breeding programme design is provided in Figure 1. As no GIFT samples from the parental seventh generation were available for analysis, GIFT broodstock (n=50) from the nineteenth generation (WorldFish Center, Malaysia) were substituted as the closest available material to the parental GIFT stock (Supplementary Table S1).

The HotSHOT method was used to prepare crude genomic DNA of the GIFT fin clip samples (Truett et al., 2000). Purified DNA was extracted by a modified salt precipitation method (Syaifudin et al., 2019). Small pieces of fin tissue were digested in 300 µL SSTNE lysis solution (0.3 M NaCl, 50 mM Tris base, 0.2 mM EDTA pH 8.0, 0.2 mM EGTA, 0.5 mM spermidine, 0.25 mM spermine and 0.1% SDS) containing 1.5 µL proteinase K (10 mg/mL) at 55 °C overnight. Lysed samples were treated with 5 µL RNaseA (2 mg/mL) at 37 °C for 1 h and the supernatant centrifuged twice at 21,000 ×*g* after precipitation with 180 µL 5 M NaCl on ice. The resulting DNA was precipitated in an equal volume of isopropanol, washed twice in 70% ethanol and dissolved in TE buffer (10 mM

Tris, 1 mM EDTA pH 8.0) until DNA quantification. The quantity and quality of DNA were assessed by measurement on a Nanodrop spectrophotometer (Labtech International Ltd, UK) and by agarose gel electrophoresis. Standardised dilutions of 8 ng/μL DNA were prepared in 5 mM Tris buffer pH 8.0.

2.3 Species-specific diagnostic SNP markers

Ten SNP markers were selected for this study (Syaifudin et al., 2019) based on the ability to clearly distinguish among three species (four with an allele specific for *O. niloticus*, four for *O. mossambicus* and two for *O. aureus*) as indicated by a high frequency of the diagnostic allele (97% for one of the *O. niloticus* markers, 100% for the other nine) in the target species and absence of this allele in the other two species based on a test panel of 75 individuals from the three species. Details of the PCR primers are provided in Supplementary Table S2.

2.4 PCR-based SNP genotyping

Individuals were genotyped using KASP (Kompetitive Allele Specific end-point PCR) technology by LGC Genomics Ltd (UK) as detailed previously (Syaifudin et al., 2019). KASP primers were designed, manufactured and supplied at a proprietary concentration by LGC. Either 1 μL HotSHOT preparation or 8 ng of purified DNA template for each assay was dried in a single well of a 96 well white PCR plate (Starlab, UK). The PCR was conducted in a 5 μL total volume with 0.07 μL allele-specific primers in the propriety KASP Master Mix. PCR cycling conditions (TAdvanced thermocycler, Biometra) included an initial denaturation step at 94 °C for 15 min, 10 cycles at 94 °C for 20 s and touchdown 65 °C to 57 °C (dropping 0.8 °C each cycle) for 1 min followed by 35 cycles of amplification at 94 °C for 20 s and 57 °C for 1 min. Fluorescence signals were measured at 22 °C using a Quantica® Real Time PCR Thermal Cycler (Techne) and genotypes assigned by allelic discrimination analysis using the Quansoft software v1.121.

2.5 Statistical analysis

Deviation of allele frequency (Chi-square goodness of fit test; Power and Sokal, 2011) in the G7 hybrid samples from the expected 1:3 ratio (*O. niloticus*: *O. mossambicus*) in the backcross base population was calculated using an online tool [http://www.socscistatistics.com/tests/goodnessoffit/Default2.aspx]. Principal

Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC) was carried out on these SNP data using R v3.3.2 (R Core Team, 2019) and an associated R/*adegenet* package v1.4-1 (Jombart, 2008) to model the total variation within the dataset and identify clusters of genetically related individuals within the Molobicus breeding programme.

3. Results

The species-specific SNP assays allowed the species contribution of the samples from the Molobicus breeding programme to be assessed. Individual SNP genotypes of the GIFT strains and Molobicus samples for each of the ten markers tested by KASP are listed in Supplementary Table S3. A summary of the genotype distribution and allele frequency of the GIFT, parent and Molobicus strains is presented in Table 1. Among the total fish genotyped for two *O. aureus*–specific markers (n=151), only one copy of an *O. aureus* – specific allele was detected (a single heterozygote for *Oau*966 in the GIFT population). It was therefore concluded that *O. aureus* contribution to the Molobicus samples tested was negligible, signifying the alternate allele for *O. niloticus*-specific markers indicated an *O. mossambicus* allele and vice versa (*i.e.* the alternate allele for *O. mossambicus*–specific markers indicated an *O. niloticus* allele).

3.1 GIFT tilapia

The 50 GIFT tilapia samples (nineteenth generation) were found to possess predominantly the diagnostic allele at the *O. niloticus*-specific SNPs and the alternate allele at the *O. mossambicus*-specific SNPs (combined mean of 0.88 *O. niloticus* allele frequency and 0.12 *O. mossambicus* allele frequency, based on the previous assumption that only two species contributed).

3.2 Feral *O. mossambicus* tilapia (parent of Molobicus hybrid)

The majority of the feral *O. mossambicus* (n=22) used as parents for the Molobicus hybrid programme were noted to be homozygous for the diagnostic allele at all four of the *O. mossambicus*-specific SNP markers, and homozygous for the alternate allele at all four *O. niloticus*-specific markers (combined mean of 0.98 *O. mossambicus* alleles, 0.02 *O. niloticus* alleles). The one exception, sample MoMo-14-1, presented with a heterozygous genotype for three out of the four *O. mossambicus* SNP markers selected. The corresponding allele frequency for *O. niloticus* specific markers was low (mean 0.03), with the diagnostic marker only evident in two individuals as a heterozygous genotype.

3.3 F1 parental cross (GIFT × feral *O. mossambicus*)

Aside from *Oni*3057 (five homozygotes present for the alternate allele) and

*Omo*2007 (one homozygote for the diagnostic allele), all of the F_1 fish were found to be heterozygous for every marker tested (excluding the *O. aureus*-specific markers). Overall, the mean diagnostic allele frequency for the eight diagnostic *O. niloticus* and *O. mossambicus* markers was 0.47 and 0.51 respectively, close to the expected 1:1 ratio (P=0.55 and P=0.84, respectively).

3.4 Seventh generation (G7) Molobicus hybrid tilapia

Given the genotyping results of the parental fish, the backcross base population was predicted to have a 1:3 (*O. niloticus*:*O. mossambicus*) allelic ratio for the eight SNP markers that distinguished between these two species, with the possible exception of *Oni*3057 (due to the 38:63 ratio, of the respective diagnostic and alternate allele in the F_1 samples). Analysis of the genotyping results for the seventh generation (G7) Molobicus hybrid samples in this study (n=58, both lines combined) indicated a significant (P<0.01) deviation from this 1:3 ratio, favouring *O. niloticus*-associated alleles at the expense of the *O. mossambicus*-associated alleles, for six out of the eight SNPs, while the two remaining markers (*Oni*3057 and *Omo*2007) did not deviate from this ratio (data not shown). When the two lines were analysed separately, three of the eight loci showed a significant excess (P<0.01) of *O. niloticus*-associated alleles in the line selected in the extensive pond culture system (n=24), while in the intensive farming system population (n=34), a significant excess (P<0.01) of *O. niloticus*-associated alleles was noted in five out of the eight studied loci and in two further loci at a lower significance level (P<0.05; Table 1).

3.5 Discriminant Analysis of Principal Components (DAPC)

Discriminant Analysis of Principal Components (DAPC) was conducted using R/*adegenet* based on the ten SNP markers. The dataset included the 151 samples from the present study and the genotypes of 60 individuals representing the three pure tilapia species of *O. niloticus*, *O. mossambicus* and *O. aureus* (Syaifudin et al., 2019) as the reference populations. DAPC analysis was able to clearly separate the pure species of *O. niloticus* (coloured dark orange), *O. mossambicus* (dark blue) and *O. aureus* (green) into three distinct groups using both component comparisons (Figure 2). From the current study, the GIFT samples genotyped (light orange) were positioned as a broad cluster closer to *O. niloticus* than to

O. mossambicus. The feral *O. mossambicus* (light blue), although with an elliptical distribution, overlapped the reference *O. mossambicus* strains. In contrast, the F_1 Molobicus hybrid strains (coloured pink) were placed at an intermediate distance between the pure *O. niloticus* and *O. mossambicus* populations. Following selection, the G7 Molobicus hybrid strains formed two overlapping clusters extending beyond the F_1 hybrids. The discriminant analysis supported the closer association of the G7 Molobicus hybrid strains farmed in the extensive culture system (light grey) relative to the feral *O. mossambicus* parental strains compared to the group reared in the intensive culture system (dark grey) that were positioned more adjacent to the clusters containing the *O. niloticus* pure species and GIFT population.

4. Discussion

4.1 SNP methodology

The identification of tilapia species is of importance for the management of farmed and wild stocks due to the existence of multiple tilapia species and potential hybrids. A panel of ten validated species-diagnostic SNP assays exploiting KASP technology were applied to confirm the genotype of individuals from the Molobicus breeding programme. The number of SNPs applied was a compromise based on the expected species contribution, information gained and minimising the cost within a large-scale breeding programme.

KASP technology was chosen for the SNP assays due to the flexibility, low cost and ease of use compared to array-based platforms (Semagn et al., 2014). The HotShot DNA extraction method was able to generate crude DNA template suitable for the PCR assay and offer equivalent performance (data not shown) and a rapid alternative to the longer salt precipitation protocol yielding purified DNA. The accuracy of KASP genotype call was confirmed in the original validation study by the high level of agreement 99.4% found between the PCR based assay and ddRADseq data for the panel of 24 species specific SNPs and 34 tilapia samples (Syaifudin et al., 2019), with disagreement noted by the inherent bias towards homozygotes in the RADseq method (Davey et al., 2013). It is therefore recommended that SNP genotypes derived from KASP assays rather than RADseq studies be relied upon for small scale SNP profiling due to the improved accuracy especially, as in this study, when heterozygotes are anticipated.

4.2 Interpretation of genotype and species contribution

The expectation was that the species contribution involved in the Molobicus hybrid would be primarily from *O. mossambicus* and *O. niloticus*, but two *O. aureus*-specific markers were included as *O. aureus* is another tilapia species that has been widely transferred through aquaculture. Apart from a single copy of the *O. aureus* diagnostic allele in a single GIFT individual, the data did not show any evidence of *O. aureus* contribution to the Molobicus hybrid, so it was assumed that the alternate allele for the *O. niloticus*-specific markers indicated an allele of *O. mossambicus* origin, and vice-versa.

4.3 GIFT broodstock genotype

Based on the selected SNP panel, the KASP results suggested that individuals from the GIFT broodstock population (nineteenth generation) were mainly composed of the *O. niloticus* species (mean 0.82 diagnostic allele frequency) with a minor contribution from *O. mossambicus* (mean 0.07) and negligible contribution from *O. aureus*. These results are in agreement with previous SNP genotyping studies that noted a close association between GIFT and *O. niloticus* individuals (Van Bers et al., 2012; Xia et al., 2015), also reinforced by the close placement of the GIFT population to the reference *O. niloticus* samples following DAPC analysis within this study.

On the basis of the SNP markers analysed in the present study, and the assumption (explained above) that only *O. niloticus* and *O. mossambicus* contributed to GIFT, it appears that the nuclear genome of GIFT is around 88% *O. niloticus* and 12% *O. mossambicus*, but with only eight markers and the frequencies per locus ranging from 63 to 100% *O. niloticus*-specific alleles, plus the evidence for selection affecting most of these markers in Molobicus, this is only an estimate. Evidence of introgression by *O. mossambicus* within the GIFT strain has been documented before and the most likely source of *O. mossambicus* introgression would have arisen from the Asian farmed stocks used in the GIFT base population (Taniguchi et al., 1985). This minor level of introgression by *O. mossambicus* is consistent with previous genotyping studies suggestive of a lower than 20% *O. mossambicus* admixture in certain GIFT individuals when assignment testing was applied to estimate the genetic structure of GIFT samples based on combined mtDNA and microsatellite data (McKinna et al., 2010), microsatellite data (Sukmanomon et al., 2012) and SNP sequencing (Xia et al., 2014), but not when mtDNA haplotypes were considered alone (40% O. mossambicus mtDNA; McKinna et al., 2010).

McKinna et al. (2010) also concluded that 2 of 30 GIFT tilapia analysed (7%) had *O. aureus* mitochondrial DNA haplotypes, but it seems likely that these originated from West African *O. niloticus*, which have mtDNA haplotypes typical of *O. aureus* despite having nuclear genomes related to *O. niloticus* (Rognon and Guyomard, 2003; Syaifudin et al., 2019). This current study supports a lower

(negligible) contribution of *O. aureus* within GIFT tilapia. Previous genotyping studies have also noted GIFT individuals that contained trace levels of genetic variation suggestive of *O. aureus* or a third species involvement other than *O. niloticus* and *O. mossambicus* in alignment with the present findings (Sukmanomon et al., 2012; Xia et al., 2014). Further discriminatory genotyping studies will be required to assess the level of possible introgression by *O. aureus* within the GIFT population.

4.4 Feral *O. mossambicus* genotype

It is perhaps surprising that the species-specific SNP profiles implied that the feral *O. mossambicus* sourced from the Philippines and used as parents for the Molobicus breeding programme had only a trace contribution from *O. niloticus* (0.03 mean allele frequency). This was largely due to two individuals that were multiple heterozygotes. Although *O. mossambicus* was the first tilapia species introduced into the country in the mid-1950s, the findings suggest that this particular population of feral *O. mossambicus* has been able to maintain a high level of genetic purity in the wild despite the later introduction of domesticated and inevitable release of feral *O. niloticus* into the same environment (Pullin et al., 1997).

4.5 F1 Molobicus hybrid stock genotype

The GIFT samples analysed were derived from a later generation (nineteenth) than used in the development of Molobicus, which may account for why the F_1 fish (seventh generation GIFT × feral *O. mossambicus*) were found to be heterozygous for the markers tested, with one exception. SNP *Oni*3057, where 0.38 *O. niloticus* diagnostic alleles were observed in the F_1 individuals, was one of the two loci with the lowest frequency of *O. niloticus* diagnostic alleles in the GIFT samples. The almost uniform observation of heterozygotes for the other seven loci in the F_1 strains led to testing the seventh generation Molobicus data against an expected 1:3 (*O. niloticus*:*O. mossambicus*) allelic ratio.

4.6 G7 Molobicus species contribution and culture system

De Verdal et al. (2014) showed that the Molobicus line selected for performance in a tank-based culture system at high stocking density, with an average salinity of 2.2 ppt, and fed *ad libitum* responded to a greater extent to selection for body

weight than the line reared in the earthen ponds at low stocking density with no external feed input and a lower salinity level of 1.5 ppt. This correlated with a greater shift towards *O. niloticus*-specific alleles in the intensive line that responded more strongly to selection (mean *O. niloticus* allele frequency 0.44 in the intensively reared line vs 0.33 in the extensively reared line, with seven significant increases in *O. niloticus* allele frequency (P<0.05) from the predicted 1:3 *O. niloticus*: *O. mossambicus* ratio compared to three respectively). Analysis of the SNP dataset by DAPC also supported these findings, shown by the relative position of the G7 population reared under intensive conditions adjacent to the *O. niloticus* species compared to the extensive G7 population cluster positioned closer to the parental feral *O. mossambicus* group.

4.7 Diagnostic SNP markers and trait association

It could be anticipated that the *O. niloticus* genome would carry more alleles for faster growth at genes affecting this trait, while *O. mossambicus* could possess more allelic variants for greater salinity tolerance at genes affecting this trait, given the known attributes of these species. Similarly, the species associated SNPs used in this study and distributed throughout the genome could reflect these and other traits particular to a species, as seen by the two basic patterns of species-specific allele frequency (1:3 or 1:1) according to the species-specific marker tested. Certainly, the average body weight of the Molobicus hybrids farmed in the intensive system at G4 was reported to be increased and growth more rapid than the low input environment (de Verdal et al., 2014). Enhanced growth is a known attribute of the *O. niloticus* species and the main selection drive within the GIFT selection programme (Ponzoni et al., 2011), so perhaps a proportion of the *O. niloticus* SNP panel could reflect a growth advantage. Efforts on ongoing to unravel the genetic basis for growth selection with polymorphisms in the growth hormone gene implicated in *O. niloticus* (Jaser et al., 2017) and multiple linkage groups (LGs) associated with growth in saline tolerant hybrid tilapia derived from *O. mossambicus* and Asian red tilapia (Lin et al., 2016).

Likewise, a subset of the SNP markers could have been influenced by the differing levels of salinity between the two culture systems, although both hybrid populations were able to tolerate the brackish conditions, outside the optimum

range (0 to 1.0 ppt) for *O. niloticus* (Villegas, 1990). However, questions could be raised as to whether the salinity was actually at a sufficient level in the Molobicus breeding programme to impose strong differential selection favouring regions of the *O. mossambicus* genome associated with salinity tolerance. Given the benefits of extending the culture of tilapia into brackish environments, different approaches have been put forward to maximise this resource in coastal regions and where water sources are limited (Cnaani and Hulata, 2011). Salt tolerance has been assessed using GIFT strains grown in seawater (Ridha, 2008) and using hybrids between *O. niloticus* and *O. mossambicus* in Thailand (Kamal and Mair, 2005). Recent studies have attempted to characterise the underlying genetic mechanisms involved in tilapia salinity tolerance and identified Prolactin I (PRL I; Streelman and Kocher, 2002; Velan et al., 2015) and the Enhancer of Polycomb Homolog 1 (EPC1) as possible candidate genes involved in osmoregulation (Gu et al., 2018; Su et al., 2020).

The advent of high throughput sequencing and completion of the genome assembly for commercially important tilapia species such as *O. niloticus* (Conte et al., 2017) should help accelerate the identification and genetic manipulation of key traits. SNP datasets are available for the three species included here and Red tilapia (Kajungiro et al., 2019; Van Bers et al., 2012; Xia et al., 2014). This technology was demonstrated in a later whole genome sequencing study able to locate the signatures of selection in multiple LGs most common in non-coding regions as well as known growth-related pathways in genetically improved tilapia lines (Xia et al., 2015). The ten diagnostic SNP markers employed in the current study also represent multiple LGs throughout the genome and are positioned in non-coding regions, however their functional significance is unknown. Further analysis of whether there have been differential changes across the Molobicus genome, in terms of the contribution of *O. mossambicus* and *O. niloticus*, would require a more detailed analysis using large SNP sets or resequencing, and to associate genomic regions with the traits under selection.

Conclusions

A set of ten species-specific SNP markers diagnostic for the commercially important tilapia species *O. niloticus*, *O. mossambicus* and *O. aureus* was applied

to hybrids involved in the Molobicus breeding programme, developed from GIFT × *O. mossambicus* crosses and selected for growth performance in brackish water. Following seven generations of selection, the SNP profiling results indicated that there had been a shift in the original species contribution within the hybrid population in favour of *O. niloticus* at the expense of *O. mossambicus* alleles and that this effect was more pronounced in the line selected in an intensive culture system, which also showed a greater response to growth selection, compared to the line selected in an extensive farming environment. This is the first case study to demonstrate the utility of species-specific SNP markers in the identification of tilapia species and assessment of changes within tilapia hybrids under selection. Equally, this discriminatory SNPs method offers a particular value in assessing the threat of hybridisation in native populations of tilapias (FAO, 2019).

Acknowledgments

This work was partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions. The authors also acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance (CSC reference BDCA-2013-5). Support was also provided by the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The program is supported by contributors to the CGIAR Trust Fund. Publication cost was covered by the University of Stirling APC fund.

Author contributions

The study was designed by HdV and DJP. WR, NM, JAHB, BJM, MS, HdV and JBT contributed samples. The laboratory work was conducted by KLB. KLB, KT, MB and SW conducted the data analysis. KLB, KT and DJP drafted the initial manuscript. All authors read, edited and approved the manuscript.

References

Angienda, P.O., Lee, H.J., Elmer, K.R., Abila, R., Waindi, E.N., Meyer, A., 2011. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. Conserv. Genet. 12, 243–255. https://doi.org/10.1007/s10592-010-0136-2

Bardakci, F., Skibinski, D.O.F., 1994. Application of the RAPD technique in tilapia fish: Species and subspecies identification. Heredity (Edinb). 73, 117–123. https://doi.org/10.1038/hdy.1994.110

Cai, W.Q., Li, S.F., Ma, J.Y., 2004. Diseases resistance of Nile tilapia (*Oreochromis niloticus*), blue tilapia (*Oreochromis aureus*) and their hybrid (female Nile tilapia x male blue tilapia) to *Aeromonas sobria*. Aquaculture 229, 79–87. https://doi.org/10.1016/S0044-8486(03)00357-0

Cnaani, A., Hulata, G., 2011. Improving salinity tolerance in tilapias: Past experience and future prospects. Isr. J. Aquac. - Bamidgeh 63, 533.

Conte, M.A., Gammerdinger, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D., 2017. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. BMC Genomics 18, 341. https://doi.org/10.1186/s12864-017-3723-5

Costa-Pierce, B.A., 2003. Rapid evolution of an established feral tilapia (*Oreochromis* spp.): the need to incorporate invasion science into regulatory structures, in: Marine Bioinvasions: Patterns, Processes and Perspectives. Springer Netherlands, Dordrecht, pp. 71–84. https://doi.org/10.1007/978-94-010-0169-4_7

D'Amato, M.E., Esterhuyse, M.M., van der Waal, B.C.W., Brink, D., Volckaert, F.A.M.M., 2007. Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping. Conserv. Genet. 8, 475–488. https://doi.org/10.1007/s10592-006-9186-x

500 Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics.
501 Brief. Funct. Genomics 9, 416–423. https://doi.org/10.1093/bfgp/elq031

502 Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., Blaxter, M.L.,
503 2013. Special features of RAD Sequencing data: Implications for
504 genotyping. Mol. Ecol. 22, 3151–3164. https://doi.org/10.1111/mec.12084

505 de Verdal, H., Rosario, W., Vandeputte, M., Muyalde, N., Morissens, P., Baroiller,
506 J.-F.F., Chevassus, B., 2014. Response to selection for growth in an
507 interspecific hybrid between *Oreochromis mossambicus* and *O. niloticus* in
508 two distinct environments. Aquaculture 430, 159–165.
509 https://doi.org/10.1016/j.aquaculture.2014.03.051

510 Dinesh, K.R.R., Lim, T.M.M., Chan, W.K.K., Phang, V.P.E.P.E., 1996. Genetic
511 variation inferred from RAPD fingerprinting in three species of tilapia.
512 Aquac. Int. 4, 19–30. https://doi.org/10.1007/BF00175218

513 Eknath, A.E., Bentsen, H.B., Gjerde, B., Tayamen, M.M., Abella, T.A., Gjedrem,
514 T., Pullin, R.S.V., 1991. Approaches to national fish breeding programs –
515 Pointers from tilapia pilot studies. NAGA, ICLARM Q. 14, 10–12.

516 Eknath, A.E., Dey, M.M., Reyes, M., Gjerde, B., Abella, T.A., Sevilleja, R., 1998.
517 Selective breeding of Nile Tilapia for Asia, in: Proceedings of the 6th World
518 Congress on Genetics Applied to Livestock Production 27. University of
519 New England, Armidale, pp. 89–96.

520 FAO, 2019. FAO Yearbook. Fishery and Aquaculture Statistics 2017. Rome.

521 Firmat, C., Alibert, P., Losseau, M., Baroiller, J.F., Schliewen, U.K., 2013.
522 Successive Invasion-Mediated Interspecific Hybridizations and Population
523 Structure in the Endangered Cichlid *Oreochromis mossambicus*. PLoS One 8,
524 e63880. https://doi.org/10.1371/journal.pone.0063880

525 Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding
526 in aquaculture to meet future demands for animal protein: A review.
527 Aquaculture 350–353, 117–129.
528 https://doi.org/10.1016/j.aquaculture.2012.04.008

529 Gu, X.H., Jiang, D.L., Huang, Y., Li, B.J., Chen, C.H., Lin, H.R., Xia, J.H., 2018.
530 Identifying a Major QTL Associated with Salinity Tolerance in Nile Tilapia
531 Using QTL-Seq. Mar. Biotechnol. 20, 98–107.
532 https://doi.org/10.1007/s10126-017-9790-4

533 Jaser, S.K.K., Dias, M.A.D., Lago, A. de A., Reis Neto, R.V., Hilsdorf, A.W.S.,
534 2017. Single nucleotide polymorphisms in the growth hormone gene of
535 *Oreochromis niloticus* and their association with growth performance.
536 Aquac. Res. 48, 5835–5845. https://doi.org/10.1111/are.13406

537 Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic
538 markers. Bioinformatics 24, 1403–1405.
539 https://doi.org/10.1093/bioinformatics/btn129

540 Kajungiro, R.A., Palaiokostas, C., Pinto, F.A.L., Mmochi, A.J., Mtolera, M.,
541 Houston, R.D., de Koning, D.J., 2019. Population Structure and Genetic
542 Diversity of Nile Tilapia (*Oreochromis niloticus*) Strains Cultured in
543 Tanzania. Front. Genet. 10, 1269. https://doi.org/10.3389/fgene.2019.01269

544 Kamal, A.H.M.M., Mair, G.C., 2005. Salinity tolerance in superior genotypes of
545 tilapia, *Oreochromis niloticus*, *Oreochromis mossambicus* and their hybrids.
546 Aquaculture 247, 189–201.
547 https://doi.org/10.1016/j.aquaculture.2005.02.008

548 Kumar, S., Banks, T.W., Cloutier, S., 2012. SNP discovery through next-
549 generation sequencing and its applications. Int. J. Plant Genomics 2012, 1–
550 15. https://doi.org/10.1155/2012/831460

551 Lin, G., Chua, E., Orban, L., Yue, G.H., 2016. Mapping QTL for Sex and Growth
552 Traits in Salt-Tolerant Tilapia (*Oreochromis* spp. X *O. mossambicus*). PLoS
553 One 11, e0166723. https://doi.org/10.1371/journal.pone.0166723

554 Macaranas, J.M., Taniguchi, N., PANTE, M.J.R.R., CAPILI, J.B., Pullin, R.S.V.
555 V., 1986. Electrophoretic evidence for extensive hybrid gene introgression
556 into commercial *Oreochromis niloticus* (L.) stocks in the Philippines. Aquac.
557 Res. 17, 249–258. https://doi.org/10.1111/j.1365-2109.1986.tb00111.x

558 Mateo, D., Aguilar, R., Campos, W., Katalbas, M.S.F., Sanares, R., Edra, R.,

559 Chevassus, B., Lazard, J., Morisens, P., Baroiller, J.F., Rogñon, X., 2004.
560 Salinity tolerance of *Oreochromis niloticus* and *O . mossambicus* F1 hybrids
561 and their successive backcross, in: Bolivar, R., Mair, G., Fitzsimmons, K.
562 (Eds.), 6th International Symposium on Tilapia in Aquaculture. Bureau of
563 Fisheries and Aquatic Resources, Manilla, pp. 426–438.

564 McKinna, E.M., Nandlal, S., Mather, P.B., Hurwood, D.A., 2010. An
565 investigation of the possible causes for the loss of productivity in genetically
566 improved farmed tilapia strain in Fiji: Inbreeding versus wild stock
567 introgression. Aquac. Res. 41, e730–e742. https://doi.org/10.1111/j.1365-
568 2109.2010.02539.x

569 Modadugu, V.G., Acosta, B., 2004. A review of global tilapia farming practices.
570 Aquac. Asia 9, 7–16.

571 Neira, R., 2010. Breeding in Aquaculture Species: Genetic Improvement
572 Programs in Developing Countries, in: 9th World Congress on Genetics
573 Applied to Livestock Production.

574 Palaiokostas, C., Bekaert, M., Khan, M.G.Q.Q., Taggart, J.B., Gharbi, K.,
575 McAndrew, B.J., Penman, D.J., 2013. Mapping and Validation of the Major
576 Sex-Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) Using
577 RAD Sequencing. PLoS One 8, e68389.
578 https://doi.org/10.1371/journal.pone.0068389

579 Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012.
580 Double digest RADseq: An inexpensive method for *de novo* SNP discovery
581 and genotyping in model and non-model species. PLoS One 7, e37135.
582 https://doi.org/10.1371/journal.pone.0037135

583 Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Hamzah, A., Bakar, K.R.A., Yee,
584 H.Y., 2011. Genetic improvement of Nile tilapia (*Oreochromis niloticus*)
585 with special reference to the work conducted by the World Fish Center with
586 the GIFT strain. Rev. Aquac. 3, 27–41. https://doi.org/10.1111/j.1753-
587 5131.2010.01041.x

588 Power, D.M., Sokal, R.R., 2011. Biometry. The Principles and Practice of

589 Statistics in Biological Research, 4th ed. W. H. Freeman.

590 Pullin, R.S. V, Capili, J.B., 1988. Genetic improvement of tilapias: problems and
591 prospects, in: Pullin, R.S.V., Bhukaswan, T., Tonguthai, K., Maclean, J.L.
592 (Eds.), The Second International Symposium on Tilapia in Aquaculture.
593 ICLARM Conf. Proc. 15. pp. 259–266.

594 Pullin, R.S. V, Palomares, M.L., Casal, C. V, Dey, M.M., Pauly, D., 1997.
595 Environmental Impacts of Tilapias, in: Fitzsimmons, K. (Ed.), Proceedings
596 of the Fourth International Symposium on Tilapia in Aquaculture. Ithaca;
597 NY, pp. 554–570.

598 R Core Team, 2019. R: A Language and Environment for Statistical Computing. R
599 Found. Stat. Comput.

600 Ridha, M.T., 2008. Preliminary observation on salinity tolerance of three sizes of
601 the GIFT and non-improved strains of the Nile tilapia *Oreochromis niloticus*.
602 Eur. J. Sci. Res. 24, 373–377.

603 Rognon, X., Guyomard, R., 2003. Large extent of mitochondrial DNA transfer
604 from *Oreochromis aureus* to *O. niloticus* in West Africa. Mol. Ecol. 12, 435–
605 445. https://doi.org/10.1046/j.1365-294X.2003.01739.x

606 Semagn, K., Babu, R., Hearne, S., Olsen, M., 2014. Single nucleotide
607 polymorphism genotyping using Kompetitive Allele Specific PCR (KASP):
608 Overview of the technology and its application in crop improvement. Mol.
609 Breed. 33, 1–14. https://doi.org/10.1007/s11032-013-9917-x

610 Sodsuk, P., McAndrew, B.J., 1991. Molecular systematics of three tilapüne genera
611 *Tilapia*, *Sarotherodon* and *Oreochromis* using allozyme data. J. Fish Biol. 39,
612 301–308. https://doi.org/10.1111/j.1095-8649.1991.tb05093.x

613 Streelman, J.T., Kocher, T.D., 2002. Microsatellite variation associated with
614 prolactin expression and growth of salt-challenged tilapia. Physiol.
615 Genomics 9, 1–4. https://doi.org/10.1152/physiolgenomics.00105.2001

616 Su, H., Ma, D., Zhu, H., Liu, Z., Gao, F., 2020. Transcriptomic response to three
617 osmotic stresses in gills of hybrid tilapia (*Oreochromis mossambicus* female

× *O. urolepis hornorum* male). BMC Genomics 21, 110. https://doi.org/10.1186/s12864-020-6512-5

Sukmanomon, S., Kamonrat, W., Poompuang, S., Nguyen, T.T.T.T., Bartley, D.M., May, B., Na-Nakorn, U., 2012. Genetic changes, intra- and inter-specific introgression in farmed Nile tilapia (*Oreochromis niloticus*) in Thailand. Aquaculture 324–325, 44–54. https://doi.org/10.1016/j.aquaculture.2011.10.025

Syaifudin, M., Bekaert, M., Taggart, J.B., Bartie, K.L., Wehner, S., Palaiokostas, C., Khan, M.G.Q.Q., Selly, S.-L.L.C., Hulata, G., D'Cotta, H., Baroiller, J.-F.F., McAndrew, B.J., Penman, D.J., 2019. Species-Specific Marker Discovery in Tilapia. Sci. Rep. 9, 13001. https://doi.org/10.1038/s41598-019-48339-2

Taniguchi, N., Macaranas, J.M., Pullin, R.S.V., 1985. Introgressive Hybridization in Cultured Tilapia Stocks in the Philippines. Nippon Suisan Gakkaishi (Japanese Ed. 51, 1219–1224. https://doi.org/10.2331/suisan.51.1219

Toniato, J., Penman, D.J., Martins, C., 2010. Discrimination of tilapia species of the genera *Oreochromis*, *Tilapia* and *Sarotherodon* by PCR-RFLP of 5S rDNA. Aquac. Res. 41, 934–938. https://doi.org/10.1111/j.1365-2109.2009.02366.x

Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A., Warman, M.L., 2000. Preparation of PCR-quality mouse genomic dna with hot sodium hydroxide and tris (HotSHOT). Biotechniques 29, 52–54. https://doi.org/10.2144/00291bm09

Van Bers, N.E.M.M., Crooijmans, R.P.M.A.M.A., Groenen, M.A.M.M., Dibbits, B.W., Komen, J., 2012. SNP marker detection and genotyping in tilapia. Mol. Ecol. Resour. 12, 932–941. https://doi.org/10.1111/j.1755-0998.2012.03144.x

Velan, A., Hulata, G., Ron, M., Slosman, T., Shirak, A., Cnaani, A., 2015. Association between polymorphism in the Prolactin I promoter and growth of tilapia in saline-water. Aquac. Reports 1, 5–9. https://doi.org/10.1016/j.aqrep.2015.03.001

Villegas, C.T., 1990. Evaluation of the salinity tolerance of *Oreochromis mossambicus*, *O. niloticus* and their F1 hybrids. Aquaculture 85, 281–292. https://doi.org/10.1016/0044-8486(90)90027-K

Wu, L., Yang, J., 2012. Identifications of Captive and Wild Tilapia Species Existing in Hawaii by Mitochondrial DNA Control Region Sequence. PLoS One 7, e51731. https://doi.org/10.1371/journal.pone.0051731

Xia, J.H., Bai, Z., Meng, Z., Zhang, Y., Wang, L., Liu, F., Jing, W., Yi Wan, Z., Li, J., Lin, H., Hua Yue, G., 2015. Signatures of selection in tilapia revealed by whole genome resequencing. Sci. Rep. 5, 14168. https://doi.org/10.1038/srep14168

Xia, J.H., Wan, Z.Y., Ng, Z.L., Wang, L., Fu, G.H., Lin, G., Liu, F., Yue, G.H., 2014. Genome-wide discovery and *in silico* mapping of gene-associated SNPs in Nile tilapia. Aquaculture 432, 67–73. https://doi.org/10.1016/j.aquaculture.2014.04.028

Figures

Figure 1 (1 column, 85 mm x 85 mm)

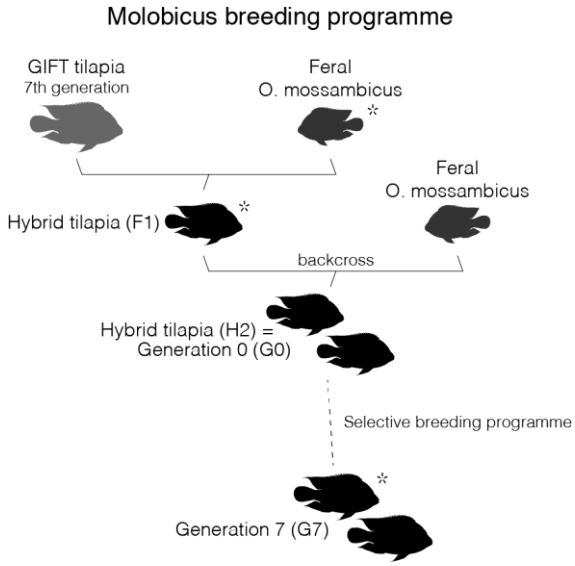

Figure 2 (2 columns, 80 mm x 167 mm)

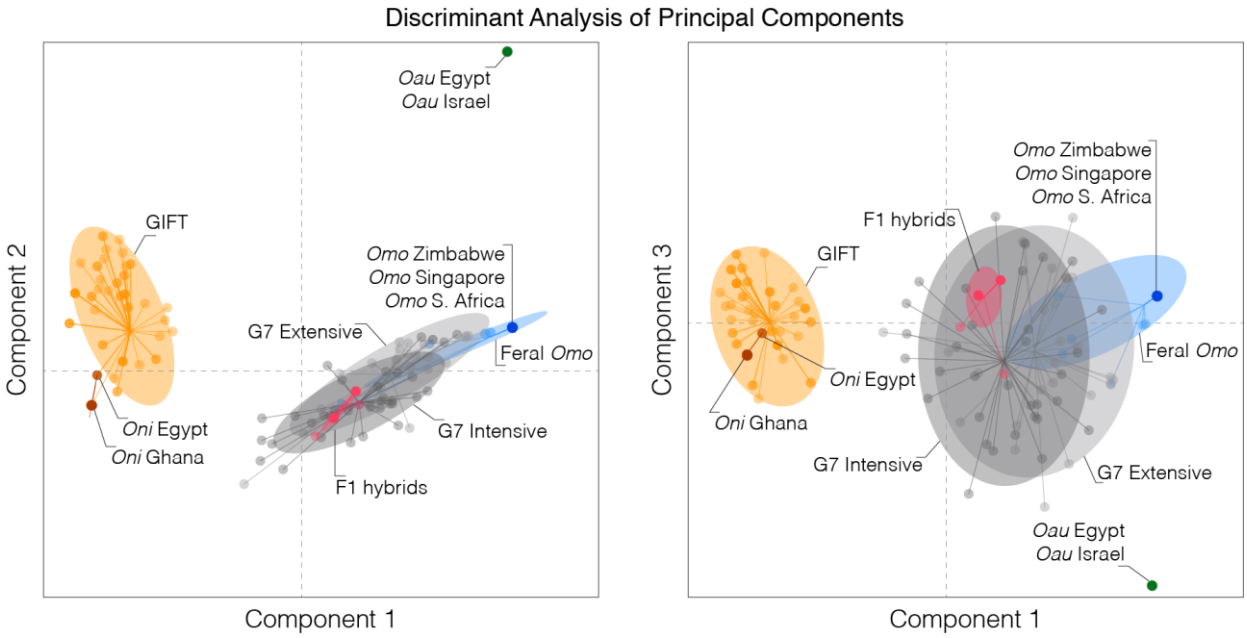

Figure Legends

Figure 1. Overview of the Molobicus breeding programme. An F_1 hybrid was initially produced from parental GIFT strains (seventh generation) and feral *O. mossambicus*, and backcrossed with *O. mossambicus*. Hybrid families underwent seven generations of selective breeding based on increased harvest weight in either extensive or intensive culture conditions. Asterisks (*) represent the three sampling points for this study.

Figure 2. Scatterplot of a Discriminant Analysis of Principal Component (DAPC), using two principal components to separate GIFT and Molobicus strains based on ten species diagnostic SNP markers of the three species *O. niloticus* (*Oni*), *O. mossambicus* (*Omo*) and *O. aureus* (*Oau*). Left panel: Component 1 and Component 2; Right panel: Component 1 and Component 3. Strains of pure tilapia species from different populations acted as a reference and were colour coded (*Oni*, dark orange, origin Egypt and Ghana; *Omo*, blue, origin Zimbabwe, Singapore and South Africa; *Oau*, green, origin Egypt and Israel), GIFT strains (light orange), feral *Omo* Molobicus parental strains (light blue), Molobicus F_1 hybrid (pink) and G7 Molobicus hybrid strains (light grey, extensive culture; dark grey, intensive culture).

Tables

Table 1. Genotype and allele frequency of Molobicus samples. Ten putative species-diagnostic SNP markers (*O. niloticus* n=4, *Oni*; *O. mossambicus* n=4, *Omo*; and *O. aureus* n=2, *Oau*) of GIFT, *O. mossambicus* parent and hybrid Molobicus strains are shown. For each marker, the number of observed genotypes (Hom, homozygous diagnostic, alternate or heterozygous) and the allele frequency (diagnostic and alternate) is reported. * $P<0.05$; ** $P<0.01$ (1:3 *Oni*:*Omo* expected ratio).

	*Oni*3057	*Oni*5782	*Oni*9497	*Oni*2675	*Omo*2007	*Omo*2657	*Omo*3481	*Omo*7956	*Oau*966	*Oau*9418
GIFT (n=50)										
Hom. diagnostic	27	50	40	20	4	0	0	0	0	0
Heterozygous	22	0	10	23	15	4	0	0	1	0
Hom. alternate	1	0	0	7	31	46	50	50	49	50
Diagnostic frequency	0.76	1.00	0.90	0.63	0.20	0.04	0.00	0.00	0.01	0.00
Alternate	0.24	0.00	0.10	0.37	0.80	0.96	1.00	1.00	0.99	1.00
O. mossambicus parents (n=23)										
Hom. diagnostic	0	0	0	0	23	22	22	22	0	0
Heterozygous	1	2	1	1	0	1	1	1	0	0
Hom. alternate	22	21	22	22	0	0	0	0	23	23
Diagnostic frequency	0.02	0.04	0.02	0.02	1.00	0.98	0.98	0.98	0.00	0.00
Alternate	0.98	0.96	0.98	0.98	0.00	0.02	0.02	0.02	1.00	1.00
F1 Molobicus hybrids (n=20)										
Hom. diagnostic	0	0	0	0	1	0	0	0	0	0
Heterozygous	15	20	20	20	19	20	20	20	0	0
Hom. alternate	5	0	0	0	0	0	0	0	20	20
Diagnostic frequency	0.38	0.50	0.50	0.50	0.53	0.50	0.50	0.50	0.00	0.00
Alternate	0.63	0.50	0.50	0.50	0.48	0.50	0.50	0.50	1.00	1.00
Molobicus G7 [extensive culture] (n=24)										
Hom. diagnostic	2	7	3	2	14	16	7	10	0	0
Heterozygous	6	13	10	6	8	6	11	8	0	0
Hom. alternate	16	4	11	16	4	2	6	6	24	24
Diagnostic frequency	0.21	0.56	0.33	0.21	0.75	0.79	0.52	0.58	0.00	0.00
Alternate	0.79	0.44**	0.67	0.79	0.25	0.21	0.48**	0.42**	1.00	1.00
Molobicus G7 [intensive culture] (n=34)										
Hom. diagnostic	1	13	4	7	24	8	8	13	0	0
Heterozygous	23	12	18	19	6	19	13	11	0	0
Hom. alternate	10	9	12	8	4	7	13	10	34	34
Diagnostic frequency	0.37	0.56	0.38	0.49	0.79	0.51	0.43	0.54	0.00	0.00
Alternate	0.63*	0.44**	0.62*	0.51**	0.21	0.49**	0.57**	0.46**	1.00	1.00

Supporting Information

Supplementary Table S1. Details of sample origin. For each sample, sample reference, species identification, country of collection and strain origin are provided.

Supplementary Table S2. SNP markers. Details of the PCR primers, diagnostic alleles, frequency and associated dyes for the ten SNP assays (from Syaifudin et al., 2019).

Supplementary Table S3. SNP markers genotypes. Lists the genotypes of each sample for the ten SNP assays.